IN THE SPECIFICATION

Please replace the paragraph at page 1, line 21 to page 2, line 3, with the following rewritten paragraph:

As this figure shows, a plurality of (in this case, N) signal sources 701 emit source signals s_i (i=1,...,N) which are mixed together and observed with a plurality of (in this case, M) sensors 702, and under these conditions the separated signals y_k (k=1,...,N) estimated to correspond to the source signals are extracted from these observed signals x_j (j=1,...,M). Here, the process that takes place between the ~~mixing of~~ source signals s_i emitted from signals sources 701 and the observations ~~of these signals~~ by sensors 702 is referred to as the "mixing process", and the process whereby the separated signals are extracted from the observations of sensors 702 is called the "separation process".

Please replace the paragraph at page 3, lines 6-19, with the following rewritten paragraph:

Since convolutional mixing problems are complicated to address and the assumption of sparsity holds better in the time-frequency domain, an effective way of addressing the above problem involves first applying a short-time discrete Fourier transform (DFT) to the abovementioned Formula (1) to transform the signal into the time-frequency domain. In the time-frequency domain, the abovementioned Formula (1) becomes

$$X(f,m) = H(f)S(f,m)$$

where f is the frequency, and m represents the timing of the DFT frames. H(f) is an (MxN) matrix whose [[ij]] <u>ji</u> element is the frequency response $H_{ji}(f)$ from signal source i to sensor j, and is referred to as the mixing matrix. Also, $S(f,m)=[S_1(f,m),...,S_N(f,m)]^T$ and $X(f,m)=[X_1(f,m),...,X_M(f,m)]^T$ are the DFT results obtained for the source signals and

observed signals, respectively. Here, the notation [α]T denotes the transposed matrix of α. Furthermore, S(f,m) and X(f,m) are vectors.

Please replace the paragraph at page 7, lines 2-4, with the following rewritten paragraph:

In cases where the number of signal sources N and the number of sensors M obey the relationship ~~M≤N~~ M<N, separation can be achieved by methods based on the signal sparsity (e.g., Non-Patent Reference 3).

Please replace the paragraph at page 7, lines 16-27, with the following rewritten paragraph:

The following method is generally used to estimate the signal source at each timing. If each signal source is assumed to be spatially separate, then between the signals observed by the plurality of sensors there will exist phase ~~shifts~~ difference and amplitude ratios determined by the relative positions of the signal sources and sensors. From the assumption that there is at most one signal contained in the observed signal at each timing, the phase differences and amplitude ratios of the observed signal at this timing correspond to the phase difference and amplitude ratio of the one signal contained in the observed signal at this timing. Accordingly, the phase differences and amplitude ratios of the observed signal in each sample can be subjected to a clustering process, and we can estimate each source signal by reconstituting the signals belonging to each cluster.

Please replace the paragraph at page 8, lines 11-19, with the following rewritten paragraph:

[0013] Next, the distribution of the relative values z(f,m) is checked and clustered into N clusters by clustering unit 752. An example of such a distribution is shown in FIG.29. In this example, a mixed signal comprising three signals (N=3) is observed by sensor 1 (j=1) and sensor 2 (j=2) — FIG.29A shows the distribution obtained using the phase difference ~~or amplitude ratio~~ alone, and FIG.29B shows the distribution obtained using both the phase difference and the amplitude ratio. As this figure shows, sparsity allows these distributions to be classified into N=3 clusters 801–803 or 811–813.

Please replace the paragraph at page 14, line 21 to page 15, line 2, with the following rewritten paragraph:

[0020] To separate limited signals consisting of signals emitted from V signal sources extracted in this way <u>in which less samples are zero padded</u>, it is possible to employ methods such as an independent component analysis method or a method in which the mixing matrix is estimated based on sparsity, for example. Consequently, it is possible to extract source signals with high quality even in cases where N>M. However, with this approach alone it is only possible to extract V source signals. Therefore, all the source signals are extracted by, for example, repeating the same processing while using a plurality of different types of mask to change the combinations of extracted signals.

Please replace the paragraph at page 15, line 20 to page 16, line 5, with the following rewritten paragraph:

[0022] To separate limited signals consisting of signals emitted from V signal sources extracted in this way <u>in which less samples are zero-padded</u>, it is possible to employ a

method such as independent component analysis or a method in which the mixing matrix is estimated based on sparsity, for example. Consequently, it is possible to extract source signals with high quality even in cases where N>M. However, with this approach alone it is only possible to extract V source signals. Therefore, for example, the same processing is repeated while using a plurality of different types of mask on a plurality of different types of set Gk to change the combinations of extracted signals. In this way, all the source signals are extracted.

Please delete Formula 9 at page 25, lines 1-2 in its entirety and replace with the following new Formula 9:

FORMULA 9

$$z_3(f,m) = \cos^{-1} \frac{z_1(f,m)v_e}{2\pi f d}$$

Please replace the paragraph at page 25, lines 14-23, with the following rewritten paragraph:

As illustrated in this figure, this histogram consists of a distribution with N (=3) peaks. In this example, clustering unit 32 clusters this distribution into N (=3) clusters (clusters 91–93 in this example). This could, for example, be performed by clustering based on a suitable threshold value, or by using methods described in many textbooks such as the k-means method or hierarchical clustering — see, e.g., ~~Morio Onoe (trans.): "Pattern Classification," Shingijutsu Communications, ISBN 4-915851-24-9,~~ <u>Richard O. Duda, "Pattern Classification (2nd ed.)", John Wiley & Sons, Inc., ISBN 0-471-05669-3,</u> chapter 10. Here, each of the resulting clusters C_i (i=1,2,...,N) is a set of relative values $z_3(f,m)$, and can be expressed as $C_i(f)=\{z_3(f,m) \mid m \in T_i\}$ using the set Ti of discrete time intervals.

Please replace the paragraph at page 27, line 25 to page 28, line 15, with the following rewritten paragraph:

First, mask generation unit 51-k reads out the variables SG_k, SG_0 and SG_k^c from temporary memory unit 90. Mask generation unit 51-k then extracts any one of the elements (a representative value within the limited range) of set G_k representing variable SG_k; this element is referred to as θ_1. Mask generation unit 51-k also extracts all the elements $G_0 \cap G_k^c$ (the representative values not inside the limited range) determined by variables SG_0 and SG_k^c, and these elements are referred to as θ_i ($i=2,\ldots,N-V+1$). Mask generation unit 51-k then stores θ_1 and θ_i in temporary memory unit 90. Next, mask generation unit 51-k extracts θ_1 and θ_i from temporary memory 90, and calculates $\tau_{ji}=(d_j/\cancel{v}\underline{v_e})\cos\theta_i$ ($j=1,\ldots,N-V+1$). Mask generation unit 51-k also calculates the elements ji of a delay matrix $H_{NBF}(f)$ from the formula $H_{NBFji}(f)=\exp(j2\pi f\tau_{ji})$ and stores them in temporary memory unit 90. In these formulae, d_j is the distance between sensor 1 and sensor j ($d_1=0$), f is a frequency variable, and [[v]] $\underline{v_e}$ is the signal velocity. These parameters could, for example, be pre-stored in temporary memory unit 90 and sequentially read out for use. The above process results in the generation of an $((N-V+1)\times(N-V+1))$ delay matrix $H_{NBF}(f)$ (FIG.3: 51a-k).

Please replace the paragraph at page 28, lines 15-27, with the following rewritten paragraph:

[0044] In this embodiment, since the relative values are taken to be the arrival directions $z_3(f,m)$ of the signals obtained from the phase difference $z_1(f,m)$ between the signals observed by two sensors, the abovementioned θ_1 represents the arrival direction of a signal corresponding to a representative value inside the limited range, and θ_i represents the arrival direction of a signal corresponding to a representative value outside the limited range. These values of θ_i ($i=1,2,\ldots,N-V+1$) are defined as shown in FIG.6. First, an origin is set in the

middle of M sensors arranged on a straight line (where L_1 is the distance from the first sensor to the origin, and L_2 is the distance from the origin to the M-th sensor and $L_1=L_2$). The angle subtended between the line connecting this origin to the i-th signal source and the line connecting the origin to the first sensor [[10]] is the angle θ_i corresponding to the i-th signal source.

Please replace the paragraph at page 29, lines 8-16, with the following rewritten paragraph:

This NBF matrix W(f) is stored in temporary memory unit 90 (FIG.1). A directional characteristics calculation unit 51c-k extracts the first row elements $W_{1k}(f)$ of this NBF matrix W(f) together with the values of d_k and [[v]] $\underline{v_e}$ from temporary memory unit 90, and generates the following directional characteristics function for the case where θ is a variable expressing the arrival direction of the signal:

FORMULA 10

$$\cancel{F(f,\theta)=\sum_{K=1}^{N-V+1} W_{1k}(f)\exp(j2\pi fd_k\cos\theta/v)} \qquad (10)$$

$$\underline{F(f,\theta)=\sum_{K=1}^{N-V+1} W_{1k}(f)\exp(j2\pi fd_k\cos\theta/v_e)} \qquad (10)$$

where θ is defined in the same way as θ_i as described above.

Please delete FORMULA 13 at page 34, lines 5-6 in its entirety and replace with following new FORMULA 13:

$$\theta_q = \arccos\frac{\arg([W_{jq}^{-1}]/[W_{j'q}^{-1}])}{2\pi f v_e^{-1} d} \qquad (17)$$

Please replace the paragraph at page 34, lines 8-16, with the following rewritten paragraph:

(where [[v]] $\underline{v_e}$ is the signal velocity and d is the distance between sensor j and sensor j')

with the representative value included in set G_k indicating the variables SG_k extracted from temporary memory unit 90, and associates the representative value a_i closest to θ_q with the q-th separated signal Y_{kq} (Step S12). In other words, permutation/scaling resolution unit 62-k applies tags Π_{kq} to the separated signals Y_{kq} representing the representative values a_i (thereby associating them with these representative values).

Please replace the paragraph at page 37, line 18 to page 38, line 3, with the following rewritten paragraph:

[0064] Also, in this embodiment, in situations where N ($N \geq 2$) signals are mixed together and observed with M sensors, a smooth-profile mask is used to separate and extract the signals. Unlike the masks used in [Conventional method 2] (a binary mask with a value of 0 or 1), a mask with this smooth profile has a profile that extends smoothly at the edges. Consequently, if this smooth-profile mask is used, then even if there are two or more observed signals at the same frequency at a certain timing and the sample relative values are separated from the representative values $a_1,\ldots,a_N$ that the sample ought to correspond to, the mask for this position may have a nonzero value, and thus it is possible to extract more signals than with a binary mask whose value changes abruptly. As a result, it is possible to suppress quality degradation resulting from zero components being padded discontinuously into the separated signals.

Please delete FORMULA 15 at page 40, lines 20-21 in its entirety and replace with following new FORMULA 15:

$$z_3(f,m) = \cos^{-1}\frac{z_1(f,m)v_e}{2\pi fd}$$

Please delete FORMULA 18 at page 44, lines 13-14 in its entirety and replace with following new FORMULA 18:

$$z_3(f,m) = \cos^{-1}\frac{z_1(f,m)v_e}{2\pi fd}$$

Please replace the paragraph at page 45, lines 3-18, with the following rewritten paragraph:

First, mask generation unit 251-k generates an (N×N) delay matrix $H_{NBF}(f)$. Specifically, mask generation unit 251-k extracts one of the representative values $a_1,a_2,\ldots,a_N$ (estimated values of the arrival directions of extracted signals) stored in temporary memory unit 90 (FIG.1), which is denoted as θ_1. Mask generation unit 251-k also extracts the other N−1 representative values (the estimated values of the arrival directions of the signals that are not extracted) from temporary memory unit 90 (FIG.1), which are denoted as θ_i (i=2,...,N). These values of θ_1 and θ_i are stored in temporary memory unit 90 (FIG.1). Mask generation unit 251-k sequentially extracts θ_1 and θ_i from temporary memory unit 90, calculates $\tau_{ji}=(d_j/\cancel{v}$ $\underline{v_e})\cos\theta_i$ (j=1,...,N) and the elements at (j,i) in delay matrix $H_{NBF}(f)$ $H_{NBFji}(f)=\exp(j2\pi f\tau_{ji})$, and sequentially stores the results in temporary memory unit 90. Here, d_j is the distance between sensor 1 and sensor j (d_1=0), f is a frequency variable, and [[v]] $\underline{v_e}$ is the signal velocity. These parameters could, for example, be pre-stored in temporary memory unit 90 and

sequentially read out when required. The above process results in the generation of an (N×N) delay matrix $H_{NBF}(f)$.

Please replace the paragraph at page 45, line 19 to page 46, line 1, with the following rewritten paragraph:

[0079] Next, mask generation unit 251-k uses this delay matrix $H_{NBF}(f)$ to produce an NBF matrix W(f) with null beamformer (NBF) characteristics. This is obtained by calculating the inverse matrix W(f) of the delay matrix $H_{NBF}(f)$ using the formula $W(f)=H_{NBF}^{-1}(f)$. This NBF matrix $W(f)=H_{NBF}^{-1}(f)$ is stored in temporary memory unit 90. Then, mask generation unit 251-k sequentially extracts the first row elements $W_{1k}(f)$ of NBF matrix W(f) and the values of d_k and [[v]] $\underline{v_e}$ from temporary memory unit 90, and generates the directional characteristics function F(f,θ) shown in Formula (10) above. After that, mask generation unit 251-k uses this directional characteristics function F(f,θ) to generate a smooth-profile mask $M_{DC}(f,m)$.

Please replace the paragraph at page 46, lines 16-20, with the following rewritten paragraph:

[0081] Here, of the estimated values of the arrival directions of the N−1 signals that are to be eliminated (i.e., the N−1 representative values other than the representative value a_i to be extracted), θ_r is the one closest to the estimated value of the arrival direction of the signal that is not eliminated (the extracted representative value a_i)<u>, for example</u>.

Please delete FORMULA 20 at page 47, lines 1-4 in its entirety and replace with following new FORMULA 20:

$$M_{DC}(f,m) = \begin{cases} |F(f,\theta_1)| & z_3(f,m) \in \text{Limited signal region} \\ |F(f,\theta_r)| & z_3(f,m) \in \text{Elimination signal region} \\ F(f,z_3(f,m)) & z_3(f,m) \in \text{Transitional region} \end{cases} \qquad (20)$$

Please replace the paragraph at page 50, lines 8-18, with the following rewritten paragraph:

[0088] First, mask generation unit 300-k reads in the relative values z(f,m), clusters C_i and representative values a_i (i=1,...,N) from temporary memory unit 90 (FIG.1) (see Steps S3-5 of the first embodiment), and calculates the variance value of each cluster C_i from the following calculation:

FORMULA 22

$$\sigma^2(f)_i = (1/|C_1|)\sum_{m \in T_i}(z(f,m) - a_i(f))^2 \qquad (22)$$

where $|C_i|$ is the number of relative values z(f,m) that belong to cluster C_i. This variance value can also be obtained by using, for example, an EM algorithm (see, e.g., ~~Morio Onoe (trans.): "Pattern Classification," Shingijutsu Communications, ISBN 4-915851-24-9, chapter 10.~~ <u>Richard O. Duda, "Pattern Classification (2nd ed.)," John Wiley & Sons, Inc., ISBN 0-471-05669-3</u>) or the like, and fitting the data to a Gaussian model.

Please replace the paragraph at page 54, lines 21-25, with the following rewritten paragraph:

[0095] Furthermore, the number of relative values z(f,m) included in the range from a_{min} to a_{max} should be at least as much as ~~the number of sensors~~ 2 and no greater than <u>the number of</u>

sensors M, and should preferably be equal to the number of sensors M. As in the first embodiment, a plurality of binary masks B(f,m) are generated in this embodiment.

Please replace the paragraph at page 64, lines 14-21, with the following rewritten paragraph:

[0116] Also, as the clustering method performed by cluster generation unit 432b, it is possible to use a method described in many textbooks, such as hierarchical clustering or k-means clustering (see, e.g., ~~Morio Onoe (trans.): "Pattern Classification," Shingijutsu Communications, ISBN 4-915851-24-9,~~ Richard O. Duda, "Pattern Classification (2nd ed.)", John Wiley & Sons, Inc., ISBN O-471-05669-3, chapter 10). Note that in any clustering method, the distance between two samples X(f,m) and X'(f,m) is defined as a means of measuring the proximity between samples, and clustering is performed so that every effort is made to include samples that are close to each other in the same clusters.

Please replace the paragraph at page 66, lines 10-19, with the following rewritten paragraph:

For example, sorting unit 433b (FIG.13) might perform the following calculation using the representative vector $a_i(f)$ for each frequency f:

FORMULA 34

$$\theta_i(f) = \arccos \frac{\arg(a_{ji}(f)/a_{j'i}(f))}{2\pi f v_{\underline{e}}^{-1}\|d_j - d_{j'}\|} \qquad \cancel{(33)}\ \underline{(34)}$$

to calculate the estimated value $\theta_i(f)$ of the arrival direction of source signal i at each frequency f. Here, d_j is the position of sensor j, [[v]] $\underline{v_e}$ is the velocity of the signal, and $a_{ji}(f)$ is the i-th element of representative vector $a_j(f)$ — for the values of d_j and [[v]] $\underline{v_e}$, it is assumed that data pre-stored in temporary memory unit 90 is used, for example.

Please replace the paragraph at page 70, line 22 to page 71, line 9, with the following rewritten paragraph:

[0129] For example permutation/scaling resolution unit 62-k could compare the estimated arrival direction $\theta_q(f)$ of the signals, which are obtained using the inverse matrix of separation matrix W(f) extracted from temporary memory unit 90 (or the Moore-Penrose pseudo-inverse matrix when $N \neq M$) by the following formula:

FORMULA 36

$$\theta_q(f) = \arccos \frac{\arg([W_{jq}^{-1}(f)]/[W_{j'q}^{-1}(f)])}{2\pi f v_{\underline{e}}^{-1} \|d_j - d_{j'}\|} \qquad \cancel{(34)}\ \underline{(36)}$$

(where [[v]] $\underline{v_e}$ is the signal velocity and d_j is the position of sensor j) with the representative vector $a_p(f)$ included in set G_k indicating the variables SG_k extracted from temporary memory unit 90, and could associate the representative vector $a_p(f)$ closest to θ_q with the q-th separated signal Y_{kq} (Step S32). In other words, permutation/scaling resolution unit 62-k applies tags Π_{kq} representing the representative values a_i to the separated signals Y_{kq} (in other words associating the tags Π_{kq} with the separated signals Y_{kq}).

Please replace the paragraph at page 73, line 19 to page 74, line 6, with the following rewritten paragraph:

[0134] As a variant of this embodiment, the limited signal values could be generated directly from the formula:

Formula 38

$$\hat{X}_k(f,m) = \begin{cases} X(f,m) & \max_{a_p(f) \in G_k} D(X(f,m), a_p(f)) < \min_{a_q(f) \in G_k^c} D(X(f,m), a_q(f)) \\ 0 & \text{otherwise} \end{cases}$$

without generating a mask M(f,m). ~~For example, limited signal generation unit 450-k could judge whether or not the observed signal vectors X(f,m) satisfy the following condition:~~

~~FORMULA 39~~

$$\cancel{\max_{a_p(f)\in G_k} D(X(f,m),a_p(f)) < \min_{a_q(f)\in G_k^c} D(X(f,m),a_q(f))}$$

~~and it could extract the observed signal vectors X(f,m) judged to satisfy this condition as the signal values emitted from the signal sources.~~

Please replace the paragraph at page 74, lines 10-15, with the following rewritten paragraph:

FIG.15 is a block diagram showing an example of the configuration of a ~~brand~~ signal separation device 500 according to this embodiment. The arrows in this figure indicate the flow of data, but the flow of data into and out from control unit 521 and temporary memory unit 522 is not shown. Specifically, even when data passes through control unit 521 or temporary memory unit 522, the associated process is not shown.

Please replace the paragraph at page 80, lines 8-15, with the following rewritten paragraph:

[0148] Also, as the clustering method, it is possible to use a method described in many textbooks, such as hierarchical clustering or k means clustering (see, e.g., ~~Morio Onoe (trans.): "Pattern Classification," Shingijutsu Communications, ISBN 4-915851-24-9,~~ Richard O. Duda, "Pattern Classification (2nd ed.)," John Wiley & Sons, Inc., ISBN 0-471-05669-3, chapter 10). Note that in any clustering method, the distance between two samples X(f,m) and X'(f,m) is defined as a means of measuring the proximity between samples, and clustering is performed so that every effort is made to include samples that are close to each other in the same clusters.

Please delete FORMULA 45 at page 83, lines 11-13 in its entirety and replace with following new FORMULA 45:

$$\theta_i = \cos^{-1} \frac{\mathrm{angle}(A_{ji}(f)/A_{j'i}(f))}{2\pi f v_e^{-1} \left\| d_j - d_{j'} \right\|} \qquad (45)$$

Please replace the paragraph at page 85, line 19 to page 86, line 14, with the following rewritten paragraph:

[0158] On the other hand, when the number of sensors is insufficient (M<N), the separated signals Y(f,m) cannot be uniquely determined with respect to the estimated mixing matrix A(f) and observed signal vectors X(f,m). This is because there infinitely many values of Y(f,m) that satisfy the following relationship:

FORMULA 47

$$X(f,m) = A(f)Y(f,m) = \sum_{i=1}^{N} a_i(f) Y_i(f,m)$$ ~~(45)~~ (47)

However, with regard to the sparsity of the source signal, it is known that the most accurate separated signal components are found at solutions of Y(f,m) that minimize the L_1 norm:

FORMULA 48

$$L_1(Y(f,m)) = \sum_{i=1}^{N} | Y_i(f,m) |$$ ~~(46)~~ (48)

(Shun-ichi Amari: "How Can Humans and Machines Distinguish Visual and Auditory Signals? (Introduction)", Journal of IEICE, Vol. 87, No. 3, ~~pp.~~ p. 167, March 2004). When separation is performed using this sort of minimizing criterion, the separation matrix W(f,m) varies with time so that separation matrix generation unit 518 calculates a time-dependent separation matrix W(f,m) from the observed signal vector X(f,m) and estimated mixing matrix A(f) at each time interval m (Step S57), and separated signal generation unit 519

calculates the separated signal components $Y_1(f,m),\ldots,Y_N(f,m)$ from the formula $Y(f,m)=W(f,m)X(f,m)$ (Step S58).

Please replace the paragraph at page 87, lines 6-19, with the following rewritten paragraph:

Next, column selection unit 516 checks the value of variable k in temporary memory unit 522, and judges whether or not $k \leq M$ (Step S63). If $k \leq M$, column selection unit 516 selects q(k) such that

$$q(k) = \mathrm{argmax}_i |a_i(f)^H \cdot e| / ||a_i(f)|| \qquad (47)$$

and stores the result of this selection in temporary memory unit 522 (Step S64). Here, Formula (47) maximizes the absolute value of the dot product of the residual vector e and the length-normalized column $\cancel{|a_i(f)^H|}\underline{a_i(f)^H}/||a_i(f)||$ — in other words, it represents an operation for selecting the representative vector $a_i(f)$ closest to the direction of the residual vector e. The reason for selecting the representative vector $a_i(f)$ closest to the direction of the residual vector e is that the residual vector e becomes smaller in the next iteration, and thus each subsequent value of $Y_i(f,m)$ becomes smaller so that ultimately it can be expected that the L_1 norm of $Y(f,m)$ defined by Formula (46) also becomes smaller.

Please replace the paragraph at page 88, line 16 to page 89, line 9, with the following rewritten paragraph:

[0163] Then, at Step S63, when column selection unit 516 judges that $k \leq M$ (equivalent to the selection of min(M,N) representative vectors $a_i(f)$), column selection unit 516 ends the loop process of Steps S64–68. At this point, the M selected representative vectors $a_{q(i)}$ span the full space, so the residual vector e becomes zero. When the loop process of Steps S64–68 ends, matrix generation unit 517 reads these M selected representative vectors $a_{q(i)}$ from

temporary memory unit 522, and generates column vectors a_i'(f,m) in which the N-M representative vectors (column vectors) a_i(f) that were not selected in the processing of steps S63–68 are set to zero (Step S69):

FORMULA 49

$$a_i'(f,m) = \begin{cases} a_i(f) & i \in \{q(1),\ldots,q(M)\} \\ 0 & i \notin \{q(1),\ldots,q(M)\} \end{cases} \qquad \cancel{(48)}\ \underline{(49)}$$

Furthermore, matrix generation unit 517 calculates a matrix A'(f,m)=[a_1'(f,m),...,a_N'(f,m)] whose columns are the column vectors a_i'(f,m) of Formula (48) ~~(this matrix is equivalent to the matrix A'(f,m) whose columns are the min(M,N) selected representative vectors a_i(f) and the max (N-M,0)zero vectors),~~ and stores it in temporary memory unit 522 (Step S70). The matrix A'(f,m) calculated in this way is an NxM matrix of which N-M rows are zero vectors.

Please replace the paragraph at page 89, lines 10-16, with the following rewritten paragraph:

[0164] Separation matrix generation unit 518 reads this matrix A'(f,m) from temporary memory unit 522, and generates its Moore-Penrose pseudo-inverse matrix $A'(f,m)^+$ as separation matrix W(f,m) (Step S71). ~~This is equivalent to an N-row x M-column separation matrix W(f,m) which is the Moore-Penrose pseudo-inverse matrix of an M-row x N-column matrix where 0 or more of the N representative vectors a_i(f) have been substituted with zero vectors.~~

Please replace the paragraph at page 92, lines 4-12, with the following rewritten paragraph:

As mentioned above, when minimization is strictly performed in the generation of a separation matrix W(f,m) in cases where N>M, the computational load becomes very heavy. For example, since there are NCM ways of making M selections from N representative vectors $a_1(f),...,a_N(f)$, it would ~~strictly be necessary to sort a group of $_NC_M$ items~~ take a computational cost proportional to $_NC_M$ in order to strictly find the combination that minimizes the L_1 norm (Formula (46)). However, with the approximate solution shown in FIG.17, it is possible to make do with a lower computational load because the number of loop iterations only needs to correspond to the number of sensors M.

Please replace the paragraph at page 93, lines 11-25, with the following rewritten paragraph:

Next, matrix generation unit 517 reads in these min(M,N) representative vectors $a_{q(i)}$ from temporary memory unit 522, generates the column vectors $a_i'(f,m)$ as follows (Step S89):

FORMULA 50

$$a_i'(f,m)=\begin{cases} a(f) & i \in \{q(1),...,q(\min(M,N))\} \\ 0 & i \notin \{q(1),...,q(\min(M,N))\} \end{cases} \quad \cancel{(49)}\ \underline{(50)}$$

and generates the matrix $A'(f,m)=[a_1'(f,m),...,a_N'(f,m)]$ whose columns consist of min(M,N) representative vectors $a_i(f)$ and max(N-M,0) zero vectors (Step S90). After the resulting matrix A'(f,m) has been stored in temporary memory unit 522, it is read in by separation matrix generation unit 518, which calculates separation matrix W(f,m) as the Moore-Penrose pseudo-inverse matrix $A(f,m)^+$ thereof (equivalent to the inverse matrix ~~W^{-1}~~ $\underline{A'(f,m)^{-1}}$ when M=N) (Step S91). This is equivalent to an N-row x M-column separation matrix W(f,m)

which is the Moore-Penrose pseudo-inverse matrix of an M-row x N-column matrix where 0 or more of the N representative vectors $a_i(f)$ have been substituted with zero vectors.

Please replace the paragraph at page 94, lines 2-7, with the following rewritten paragraph:

The present invention is not limited to the embodiments mentioned above. For example, in the first through eighth embodiments, the extracted signals are combined after they have been returned to the time domain, but ~~when using a binary mask~~ it is also possible to transform the signals into the time domain after they have been combined in the frequency domain.

Please replace the paragraph at page 97, lines 14-19, with the following rewritten paragraph:

Also, instead of the Fourier transforms and inverse Fourier transforms used to transform between the time domain and frequency domain in the above embodiments, other forms of transformation can be used such as wavelet transformation, DFT filter banks, or polyphase filter banks (see, e.g., R.E. Crochiere and L.R. Rabiner: "Multirate Digital Signal Processing," Eaglewood Cliffs, NJ: ~~Printice~~ Prentice-Hall, 1983 (ISBN 0-13-605162-6)).

Please replace the paragraph at page 98, lines 22-25, with the following rewritten paragraph:

Input unit 660 is an input device such as a keyboard, mouse, joystick or the like. Interface 670 is, for example, an input/output port that is used for the input and/or output of data, and can be connected to various types of device such as sensors, communication boards, and memory devices.

Please replace the paragraph at page 100, lines 15-25, with the following rewritten paragraph:

When the processing by CPU 620 is started, CPU 620 reads out the observed signals $x_j(t)$ from data region 652 of external memory device 650, for example, and writes them to data region 632 of RAM 630, for example. CPU 620 then performs each of the abovementioned processes under the control of control unit 622 while sequentially reading out the signal separation program from the program region 631 of RAM 630 and the ~~signal separation program~~ observed signals $x_j(t)$ from data region 632. Here, for example, RAM 630 or external memory device 650 performs the functions of memory unit 2 or 501 in the first through ninth embodiments, and RAM 630 or register 623 performs the functions of temporary memory unit 90 or 522 in the first through ninth embodiments.